Yifeng Ding

Tsinghua University, Beijing, P.R. China | **Phone:** (+86) 18012357727 | **Email:** yifeng6@illinois.edu

EDUCATION BACKGROUND

Tsinghua University, School of Software *Aug 2018 - July 2022 (expected)*

B.S. in Software Engineering

- Overall GPA: 3.61/4.00 (Rank: 33/92)
- Selected awards: *Research Excellence Scholarship* (2021); *Academic Excellence Scholarship* (2019)

Double Major: Business Administration (For Second Bachelor Degree)

- Overall GPA: 3.80/4.00

RESEARCH INTEREST

Software engineering for artificial intelligence (SE4AI); Deep learning system testing; Software testing

PUBLICATIONS

[1] Quan Zhang, Yongqiang Tian, **Yifeng Ding**, Yu Jiang, Ting Chen, Chengnian Sun, Jiaguang Sun: "GeminiGuard: Cooperative Defense against Adversarial Attacks". Submitted to *TOSEM*, *under review*.

[2] Quan Zhang, **Yifeng Ding**, Yongqiang Tian, Jianmin Guo, Min Yuan, Yu Jiang: "AdvDoor: Adversarial Backdoor Attack of Deep Learning System". ACM SIGSOFT International Symposium on Software Testing and Analysis (**ISSTA**), Denmark, 2021.

RESEARCH EXPERIENCE

Research Assistant Advisor: Prof. Yu Jiang

Tsinghua University, Software System Security Assurance Group

Research on Defensing Deep Learning System Against Adversarial Attack *Feb 2021 – Sep 2021*

- Proposed a novel defense technique for DNN models (namely GeminiGuard, submitted to *TOSEM*), which leverages its two specifically customized components, Regulator and Inspector, to cooperatively defend against diverse adversarial attacks
- **My work includes:**
  - Proposed Inspector to capture the abnormal status in DL models and detect the adversarial examples with larger distortion
  - Conducted different experiments to verify the effectiveness of GeminiGuard on different datasets and models, against five different adversarial attack methods including PGD and C&W, and under two representative threat scenarios
  - Analyzed experimentally the correlation and collaboration between Regulator and Inspector, and proved that Regulator can regulate adversarial examples with smaller distortion and Inspector can detect that with larger distortion separately

Research on Deep Learning System Backdoor Attack *Aug 2020 – Jan 2021*

- Proposed a novel backdoor attack on DL system (namely AdvDoor, published in *ISSTA'21*), which utilizes the Targeted Universal Adversarial Perturbation (TUAP) to hide the anomalies in DL models and confuse existing detection methods
- **My work includes:**
  - Compared the effectiveness of AdvDoor and patch backdoor on CIFAR-10 and GTSRB and proved that AdvDoor can achieve higher success rate and more stable predicting accuracy on random pairs no matter how different the categories are
  - Conducted experimental study to verify the ability of AdvDoor in fooling state-of-the-art backdoor detection methodology
  - Participated in the paper writing of the research work, including Related Work, Evaluation, and Discussion parts

HONORS AND AWARDS

- **Research Excellence Scholarship** (awarded to those with outstanding research work) 2021
- **Academic Excellence Scholarship** (awarded to those with outstanding academic performance) 2019
- **Top 0.005% on the National College Entrance Exam** 2018
- **National Mathematical Olympiad in Provinces, First Prize** 2017

ADDITIONAL INFORMATION

Extracurricular Experiences

Student Science and Technology Association for School of Software, *Executive Committee* *Feb 2020 – Dec 2020*

- Planned and organized several student activities, including dissertation defense meetings of Challenge Cup competition, Student Scientific Research Training (SSRT) program and the school-wide Challenge Cup Exhibition

Social Practice Team, *Team Leader* *July 2019 – Aug 2019*

- Delivered lectures for high school outreach program, focusing on learning strategies and preparation for the national college entrance exam; Received Medium-sized Team Excellence Award among 342 social practice teams

Computer and Language Skills

- Language: Chinese (Native), English (Fluent, TOEFL: 108 for total and 24 for speaking)
- Coding skills: Python, C/C++, Java, JavaScript